INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **25**: 843–856 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/joc.1167

QUANTIFYING RANDOM MEASUREMENT ERRORS IN VOLUNTARY OBSERVING SHIPS' METEOROLOGICAL OBSERVATIONS

ELIZABETH C. KENT* and DAVID I. BERRY
Southampton Oceanography Centre, Southampton, UK

Received 30 April 2004
Revised 9 August 2004
Accepted 13 August 2004

ABSTRACT

Estimates of the random measurement error contained in surface meteorological observations from Voluntary Observing Ships (VOS) have been made on a 30° area grid each month for the period 1970 to 2002. Random measurement errors are calculated for all the basic meteorological variables: surface pressure, wind speed, air temperature, humidity and sea-surface temperature. The random errors vary with space and time, the quality assurance applied and the types of instrument used to make the observations. The estimates of random measurement error are compared with estimates of total observational error, which includes uncertainty due both to measurement errors and to observational sampling. In tropical regions the measurement error makes a significant contribution to the total observational error in a single observation, but in higher latitudes the sampling error can be much larger. Copyright © 2005 Royal Meteorological Society.

KEY WORDS: ICOADS; ship; random error; pressure; wind; air temperature; humidity; SST

1. INTRODUCTION

The production of error estimates to accompany climatological datasets calculated using meteorological data from merchant ships (known as Voluntary Observing Ships (VOS)) is increasingly recognized as important (e.g. Smith and Reynolds, 2004; Rayner *et al.*, 2005). Early surface flux climatologies made basic attempts to estimate the errors (e.g. Hsiung, 1986; Isemer and Hasse, 1987; Oberhuber, 1988), which were quoted as a percentage of the fluxes themselves. Significant progress was made by Gleckler and Weare (1997), who made an attempt to calculate the contribution of random and systematic errors and their correlations in each of the basic variables to the calculated fluxes. However, Gleckler and Weare (1997) did not have good quality estimates of the random errors in the variables, which they based on the literature. Kent *et al.* (1999) made estimates of the random measurement errors present in pressure, wind speed, sea-surface temperature (SST), air temperature and humidity from the Comprehensive Ocean–Atmosphere Dataset (COADS; Woodruff *et al.*, 1993, 1998). They used the semivariogram method (Morone, 1986), pioneered in marine meteorology by Lindau (1995), to make estimates of the random error using 4 months of data (January and July 1980 and 1993). Advances in computing have now allowed us to improve the random error estimates of Kent *et al.* (1999) by using all data between 1970 and 2002. This allows time series of error estimates to be presented and gives increased confidence in their maps. The new International COADS (ICOADS; Diaz *et al.*, 2002; Parker *et al.*, 2004) was also used, which contains extra data not originally available to COADS, e.g. a subset of records from the Met Office Marine Databank.

The new error estimates should inform any analysis performed using COADS or ICOADS. Without information on the random errors and sampling characteristics of the data, the significance of any results

* Correspondence to: Elizabeth C. Kent, Southampton Oceanography Centre, University of Southampton, Waterfront Campus, European Way, Southampton, SO14 3ZH, UK; e-mail: elizabeth.c.kent@soc.soton.ac.uk

cannot be properly assessed. Estimates of the random errors are important for, *inter alia*, data assimilation, variability analysis, optimal interpolation and inverse calculations.

This paper firstly introduces the data, metadata and quality assurance (Section 2). The semivariogram method and its application to the data are described in Section 3. Section 4 presents the results, giving both maps and time series of random measurement error estimates for pressure, wind speed, air temperature, specific humidity and SST. Section 4 also investigates how the errors depend on the height corrections applied, the methods of measurement used and the quality assurance criteria. Section 5 presents a discussion that includes a comparison with estimates of the combined sampling and measurement error calculated using the method of Brohan *et al*. (2003) as applied by Rayner *et al*. (2005). Conclusions are drawn in Section 6.

2. DATA

2.1. ICOADS

The ICOADS (Woodruff *et al*., 1998; Diaz *et al*., 2002) is a compilation of surface meteorological observations from VOS, buoys and fixed platforms. In this study we have used only reports from VOS made between 1970 and 2002. The accuracy of individual VOS observations is not consistently high, but they remain an important source of information over the ocean. In particular, VOS reports provide information on surface air temperature, near-surface humidity and pressure that cannot be reliably measured from satellites and provide all the variables required to calculate surface fluxes of heat and momentum. Importantly, they form a long-term data source giving information stretching back over 200 years. Obviously, the data quality will vary dramatically over this long period, and biases and random errors in the data need to be assessed so that consistent estimates of long-term climate variability and change can be made.

In this study we use both the ICOADS 'long marine reports', which are the reports from individual VOS, and 'monthly summary groups', which are monthly $2^\circ \times 2^\circ$ area monthly mean values constructed from the long marine reports.

2.2. Metadata

Some metadata are contained within ICOADS, e.g. information on the platform type, the transmission route and country of origin. In this study we are particularly interested in observation method metadata, of which there is a limited amount in ICOADS. Specifically, ICOADS contains information on the observation method for SST and wind speed. However, particularly in the early part of the period analysed, there are many reports without SST or wind speed measurement method indicators, and we would also like to have information on the methods used to measure other variables. This information is contained within the World Meteorological Organization (WMO) List of Selected, Supplementary and Auxiliary Ships (known as Publication 47; e.g. WMO, 1994). This publication contains information on the type of barometer, the type and exposure of dry bulb thermometers, the type and exposure of humidity sensors, the method of obtaining SST and some information on measurement heights. The information is listed by ship callsign and can be linked to ICOADS reports containing callsigns as described by Kent and Taylor (1997). Where two sources of information for a particular ship are available, e.g. for the SST measurement method, the metadata contained within ICOADS are preferred.

2.3. Quality assurance

ICOADS quality assurance uses 'trimming limits' to allow different levels of quality control to be applied to the data (Woodruff *et al*., 1987; Wolter, 1997). The quality assurance is based on differences from climatological 2° monthly means, and flags indicate whether the observation is within 2.8, 3.5 or 4.5 climatological standard deviations (denoted 2.8σ, 3.5σ and 4.5σ) from the climatological mean. Although the trimming quality assurance removes much of the poor quality data within ICOADS, it can perform badly in climatologically extreme months (Wolter, 1997; Kent *et al*., 1999). Applying the different trimming limits to ICOADS will obviously result in different values for the random error estimates.

3. CALCULATION OF RANDOM MEASUREMENT ERROR ESTIMATES

3.1. The semivariogram method

The semivariogram method has been used with VOS data in the past (Lindau, 1995, 2003; Kent *et al*., 1999; Kent and Challenor, in press) and is particularly useful for calculating random errors in data from ships. The method uses pairs of reports of the same variable from different ships which are made at similar times. The differences in the paired reports are due to spatial and, to a lesser extent, temporal differences in the field being measured and also to observational errors in each report. Observational errors can be due to different types of measurement method being used, to observer error, or coding and transmission errors and should be independent of ship separation. Differences due to the true spatial variability will increase with ship separation. A regression of the squared differences against separation distance will, therefore, identify the mean component of the contribution of spatial variability to the differences through the gradient and the mean contribution of measurement errors through the intercept. Lindau (1995) similarly estimated the temporal contribution to the error by regressing the differences between Ocean Weather Ship wind speed reports and nearby VOS reports on differences in reporting time. We present our error estimates in the same units as the variables themselves (i.e. as the square root of half the intercept variance; Kent *et al*., 1999).

3.2. Application to ICOADS reports

In this paper we regress binned mean-squared differences of each variable against report separation. All data pairs in each month with separations less than 300 km were binned into 20 km ranges. This is the approach taken by Kent and Challenor (2005), who argued that this method is more statistically sound than making a regression against individual estimates as in Kent *et al*. (1999). Differences between estimates made from the binned and individual regressions are, however, very small (Kent and Challenor, in press). The reports in each pair were required to be made at the same nominal reporting time. There will be a small contribution to the variability from any actual differences in the observation time. As in Kent and Challenor (in press), we have excluded observations made close to the coast and observational pairs where the two observations were in different ocean basins. Where height corrections have been made, we use the same methodology as Josey *et al*. (1999).

3.3. Total observational error

Rayner *et al*. (2005) described a technique presented by Brohan *et al*. (2003) for estimating random errors in a gridded dataset. These estimates of total observational error contain not only the contribution due to observational error (as estimated in the present study) but also an element due to the spatial and temporal sampling in the grid-box. The values of total observational error are calculated as representative of a single observation and are thus directly comparable to estimates of measurement error. Both the total observational error and the random measurement error will decrease as the number of observations in the grid-box increases. To estimate the total observational error, a time series of mean values is binned in ranges of the number of observations used to make up each mean value. The expected variation of the variance in each bin with the number of observations is modelled, allowing for any correlations between each observation contributing to the mean value (Jones *et al*., 1997). In this paper we apply the method separately to each grid-box in the ICOADS 2° area gridded monthly summaries. As the method models the expected variation of variability with the number of observations, a large number of monthly values with a range of observation densities are required. Therefore, we have used a longer analysis period for these estimates of total observational error (1950–2002) than that for the estimates of measurement error (1970–2002).

4. RESULTS: RANDOM MEASUREMENT ERROR ESTIMATES

4.1. Pressure

Figure 1(a) shows the spatial variation of the pressure error estimates calculated as described in Sections 3.1 and 3.2 using trimming limits of 3.5σ. Average random measurement error estimates have been calculated for 30° areas where 36 or more monthly error estimates could be calculated in the period 1970 to 2002. It was required that there were more than 100 pairs in each month and that the intercept was larger than the estimated uncertainty in the intercept. Although it is possible to introduce a bias into the error estimates by requiring that the intercept was larger than the uncertainty estimate, a repeat calculation without this requirement resulted in a negligible change to the data. It has been possible to calculate error estimates for most 30° areas in the Northern Hemisphere, the exceptions being those 30° regions in the Arctic, which are predominantly land, and additionally very data-sparse regions in the central tropical Pacific. Between 15 and 45 °S, estimates have been possible for the entire Atlantic and the coastal Pacific (although the estimates are likely to have been largely derived from observations in the north of this region). South of 45 °S the error estimates have only been possible near to New Zealand and South America. The coverage for these estimates is thus larger than that shown in Kent *et al.* (1999), who derived error estimates from 4 months' data only. The error estimates are typically smaller than those of Kent *et al.* (1999) and vary more smoothly, as expected from the much larger amount of data used in the present study. The maximum error estimates are found in the Arctic and in mid-latitude coastal regions; the smallest error estimates are in the tropics. The error estimates range from 1.2 to 3.9 hPa. Pressures are corrected for height at source, so no further adjustment is required.

Figure 2(a) shows how the pressure random error estimates vary over the period 1970 to 2002. Two estimates of the global error are shown: the error estimates weighted by the number of observations in each 30° region and month, and the error estimates weighted by the ocean area in each 30° region. It has only been possible to use the error estimates between 45 °S and 75 °N. Globally representative average errors have been calculated in the following way. Small temporal gaps of 1 month were filled by linear interpolation in time. Remaining missing error values in each 30° region were then replaced with a combination of the 12 month running mean and annual cycle (calculated from all available data in the region) of the error estimates for that 30° region. Any remaining missing values were in extremely data-sparse regions where a long-term mean could not be estimated and were filled by zonal linear interpolation. The resulting complete fields of error estimates were weighted to give either a value that represents the observational density or the ocean area. For an error estimate representative of the observational density, the interpolated complete fields were weighted by the monthly mean number of ICOADS VOS observations within each 30° area (note that this is not the number of observational pairs shown in Figure 1). To obtain a geographically representative error estimate, the same interpolated fields were weighted by the fraction of ocean area, calculated from the ETOPO5 data set (National Geophysical Data Center, 1988), within each 30° area.

The error estimates show a peak between 1971 and 1974, which is more pronounced when the estimates are weighted by the number of observations. For most of the period the estimates weighted by ocean area are smaller than those weighted by the number of observations. This is as a result of the lower error estimates in the relatively poorly sampled tropical regions compared with the more highly sampled coastal high northern latitudes and high-variability regions. However, towards the end of the period analysed, the number-weighted estimate drops below the area-weighted estimate. The cause is the reduction in the estimated random error for a particular very well sampled 30° area covering northern Europe and Scandinavia which dominates the number-weighted estimate. This demonstrates the difficulty in calculating a representative global number for the error. When weighting by ocean area it is necessary to fill in gaps in the error estimates and we increase the importance of some very poorly sampled regions, e.g. in the tropical Pacific. When weighting the estimates by the number of observations, the very well sampled regions dominate very strongly. As these are typically high-variability regions this method may be including an element of regional variability in the measurement error estimate (see Section 3). Thus, neither the maps of the error estimates nor the time series give a full picture of the variation in the error estimates. The maps and time series should be interpreted together, remembering the limitations of each method of calculation of the time series and also the time variation contained within the maps of error estimate.

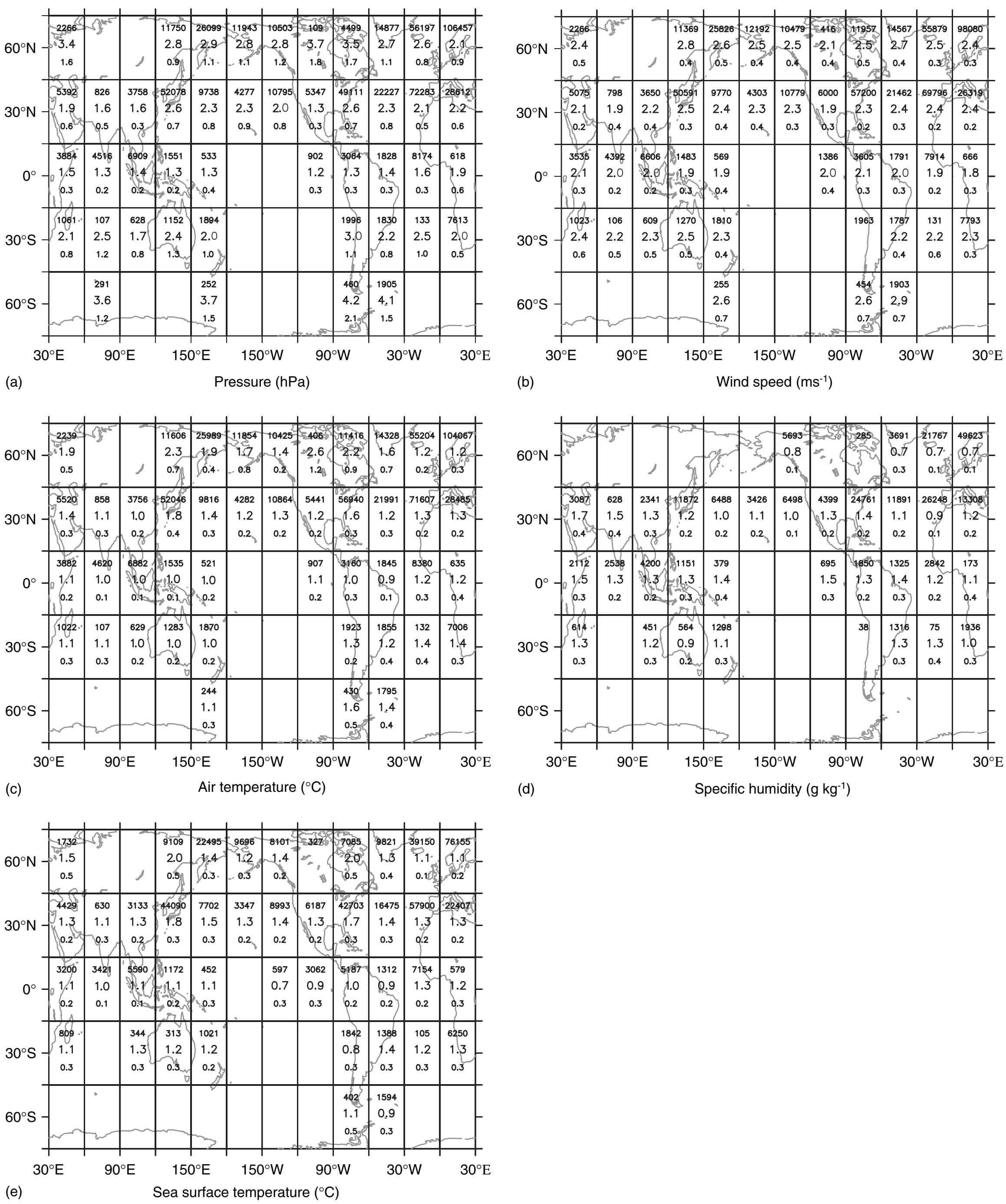

Figure 1. Random measurement error estimates for 30° area ocean regions averaged for each month in the period 1970 to 2002 calculated as described in Section 3. The large central number in each 30° region is the mean random error, the lower number is the standard deviation of the error estimates, and the upper number is the average number of observational pairs used in the calculation. Error estimates are only presented where there are more than 36 monthly estimates calculated using 100 or more data pairs

Figure 2 also shows, for error estimates weighted by the number of observations only, the error estimates unsmoothed in time. These are quite noisy, but they clearly show that the error estimates are larger in the

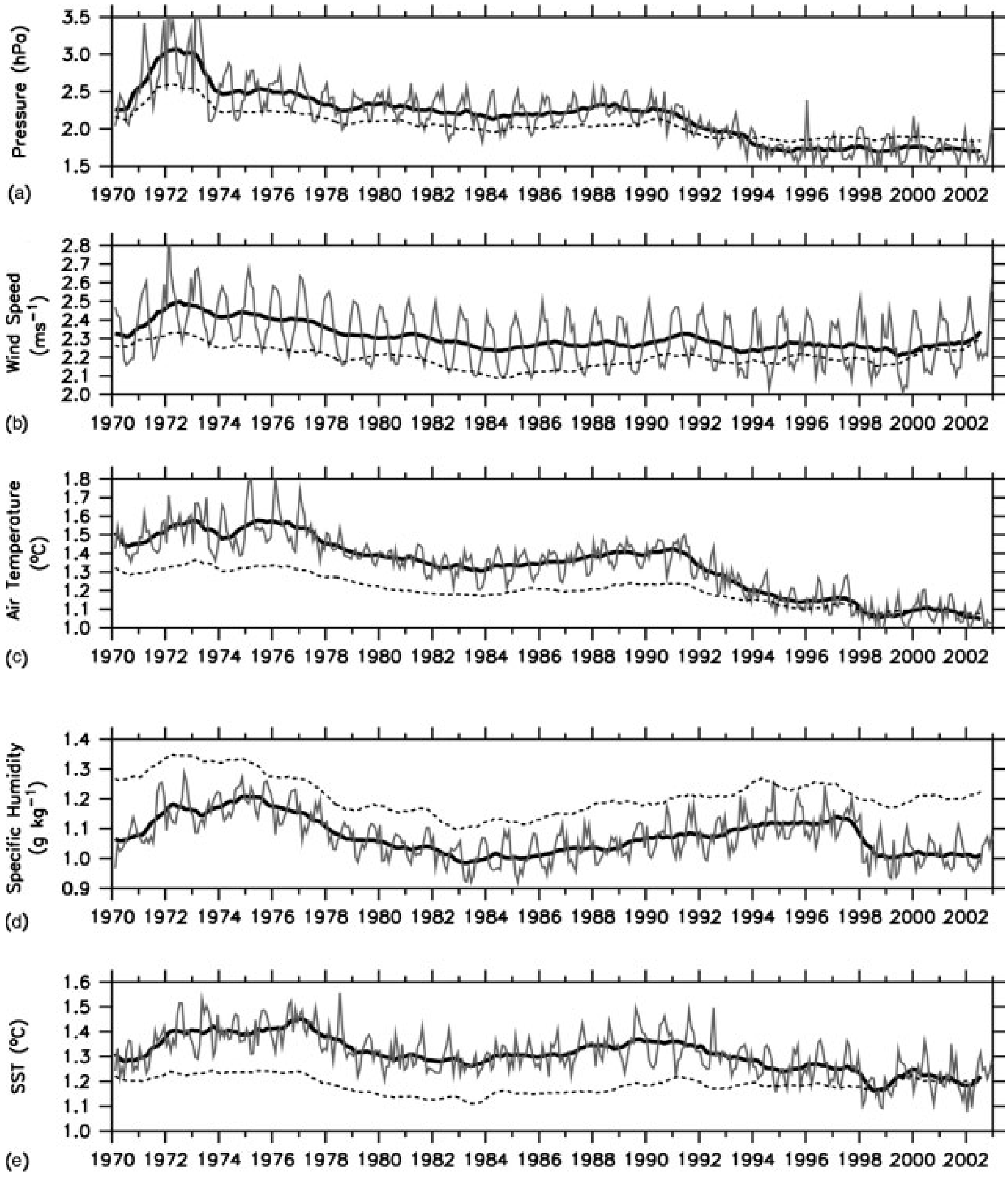

Figure 2. Time series of global average random error estimates in (a) pressure, (b) wind speed, (c) air temperature, (d) specific humidity and (e) SST (as in Figure 1). Error estimates have been smoothed with a 12 month running-mean filter and weighted by the number of observations (solid line) and by ocean area (dotted line). Also shown are the error estimates weighted by the number of observations unsmoothed in time (grey line)

Northern Hemisphere winter than in the summer. This again suggests the error estimates might be larger in regions and times of higher variability.

Within the WMO metadata there are four codes for pressure measurement method. These have been grouped together to give an analogue methods group (including the codes for aneroid barometer, ship's aneroid barometer and mercury barometer) and digital methods (comprising the code for digital aneroid barometer). Random errors were analysed separately for these two groups of instruments and it was found that for coincident estimates the ratio of random error estimates in the digital measurements to that in the analogue measurements was 0.7 ± 0.4. The digital estimates are mainly possible after the mid-1980s, as usage of digital instruments increases. The ratio increases gradually through the 1990s, suggesting that the analogue measurements are improving in quality, which is also indicated by the overall time series (Figure 2(a)).

4.2. *Wind speed*

Figure 1(b) shows the 30° area monthly random error estimates for wind speed averaged over the period 1970 to 2002 in the same format as for pressure estimates. No corrections for height or adjustments to the Beaufort scale have been applied for this calculation. This avoids the exclusion of data with an unknown measurement method. The estimates of Kent *et al.* (1999) used wind speeds corrected to 10 m neutral values and adjusted to the Beaufort Equivalent Scale of Lindau (1995). Kent *et al.* (1999) found a 15% reduction in wind speed random error estimates following height correction and conversion to the Lindau (1995) Beaufort Equivalent Scale. When this difference is taken into account, the random errors presented here are slightly smaller than those calculated by Kent *et al.* (1999). We adjusted the wind speeds for height and Beaufort Scale and recalculated the error estimates, which showed a $13 \pm 1\%$ reduction, similar to that found by Kent *et al.* (1999). It should be noted that the number of observations is reduced for this calculation, as all the variables for adjustment need to be present in each report. It was thought possible that errors in the variables used to calculate the atmospheric stability may actually act to increase the calculated random error, but this was not found to be the case; when the errors are recalculated after adjusting the wind speeds assuming neutral stability, the random error estimate is reduced by $12 \pm 1\%$. As for pressure, the random error estimate in uncorrected wind speeds is greatest in high latitudes and smallest in the tropics (Figure 1(b)). Typically, the errors in the uncorrected data are 2.2–2.3 m s^{-1}, reducing to 2.0 m s^{-1} on correction.

Figure 2(b) shows the time series of the wind speed random error estimates weighted by the number of observations and by the ocean area as in Figure 2(a). The error estimates weighted by the number of observations are larger than those weighted by the ocean area, as expected from Figure 1(b). The random errors are larger at the beginning of the period, although changes with time are small after the mid 1970s. The seasonal cycle in the wind speed error estimates is the clearest of all the variability presented and, as for the pressure error estimates, is largest in the Northern Hemisphere winter. This is consistent with the observed variability in VOS wave height measurements (Gulev *et al.*, 2003).

There is little difference between the random error estimates calculated separately for visually estimated and anemometer-measured wind speeds. The ratio of the error estimates is 1.0 ± 0.2. Kent *et al.* (1998) concluded that the errors were similar because, although the anemometer winds have the potential to be more accurate, many anemometer winds contained calibration or flow distortion errors (Yelland *et al.*, 2002; Moat *et al.*, 2005). Kent *et al.* (1998) compared satellite scatterometer winds with winds from individual ships and found that the distributions of differences for ships using anemometers tended to be narrower but could be biased, whereas the distributions for ships making visual estimates tended to be broader but less likely to be strongly biased. A bias, such as that due to a poorly calibrated anemometer, will contribute to the random error in wind speeds for a group of ships, although for the individual ship concerned it is obviously a systematic error.

4.3. *Air temperature*

Figure 1(c) shows the random error estimates for air temperatures obtained as for Figure 1(a) data. No height correction has been applied to the data. The error estimates are much more smoothly varying than those of Kent *et al.* (1999) and are typically a bit lower. This is despite the new estimates presented in Figure 1(c) being calculated from data that have not been height corrected, whereas those of Kent *et al.* (1999) were adjusted to 10 m. A recalculation of random errors for air temperature observations following height correction, including full stability effects, reduced the random error estimates shown in Figures 1(c) and 2(c) by $6 \pm 1\%$; height correction assuming neutral conditions reduced the error estimates by $4 \pm 1\%$. Although these reductions in random error are relatively small, the main purpose of the height correction is to remove bias in the data, as most air temperatures are measured at heights greater than the 10 m reference height. This is of particular concern in climate-change studies, as air temperature measurement heights have been increasing with time (Rayner *et al.*, 2003) as the ships themselves increase in size. Like pressure and wind speed, the random errors are largest in high latitudes and lowest in the tropics. Figure 2(c) shows the time variation of the random error estimates. The random errors tend to reduce with time, and those weighted by the number of observations are larger than those weighted by ocean area. The monthly variation, however, shows a different phase to pressure and wind speed, with the largest errors in the Northern Hemisphere summer.

There are significant differences between the random error estimates for air temperatures measured using different methods. Random error estimates calculated from screen-derived data are, overall, 27% greater than those derived from psychrometers (not shown), although the standard deviation of the differences is large. The differences are greatest in the early part of the period analysed, but after about 1994 the error estimates become similar. It seems probable that in recent years there has been an increased understanding of the importance of positioning the screens to ensure good exposure, perhaps prompted by the results of the VOS Special Observing Project–North Atlantic (Kent *et al.*, 1993).

4.4. *Humidity*

Figure 1(d) shows the random error estimates for specific humidity calculated without height adjustment. If the humidities are adjusted for height, the random error estimates show an $8 \pm 3\%$ decrease if the full stability-dependent correction is applied and a $7 \pm 2\%$ decrease if the correction to 10 m is calculated assuming neutral stability. The error estimates are similar in size to those of Kent *et al.* (1999). As noted by Kent *et al.* (1999), the distribution of specific humidity errors is different from that for the other variables, in that the smallest errors are found in high latitudes and the largest errors in the tropics. This is due to the variation of saturation specific humidity with temperature. The random error estimates for specific humidity are consistent with a random error in the dewpoint temperature of comparable size to that in air temperature. Figure 2(d) shows the variation with time of the specific humidity random error estimates. In this case, the ocean-area-weighted estimates are larger than those weighted by the number of observations. The most consistent humidities are reported in the early-to-mid 1980s, the largest random errors are in the early period, and there is an increase in random error towards the end of the period analysed. The error estimates are largest in the Northern Hemisphere late summer and autumn; the reason for this variation is unknown.

Random error estimates calculated from screen-derived specific humidities are 4% greater than those estimated from psychrometer reports. This ratio varies little over the analysis period. Thus, the random errors in screen and psychrometer humidities are much more similar than the air temperatures, which before 1994 showed more scatter in the screen-derived air temperatures.

4.5. *SST*

Random errors in SST have been presented and described in detail by Kent and Challenor (in press) and are included here for completeness (Figures 1(e) and 2(e)). Kent *et al.* (1999) applied corrections to SST in an attempt to account for biases in the engine intake SST (Kent *et al.*, 1993; Josey *et al.*, 1999). These corrections made little difference to the random error estimates (Kent *et al.*, 1999) and have not been applied here (nor by Kent and Challenor (in press)) as their accuracy is under review (Kent and Kaplan, accepted for publication). The SST random errors are largest in high latitudes and in high-variability regions (Figure 1(e)) and lowest in the tropics. Figure 2(e) shows the time series of random error estimates; those weighted by the number of observations are larger than those weighted by ocean area. The error estimates are largest in the Northern Hemisphere summer (Kent and Challenor, in press).

Kent and Challenor (2005) showed that random errors for engine-intake-derived SST are significantly larger than those for bucket-derived SST.

4.6. *Effect of quality assurance*

Table I gives the ratios of the errors for each basic variable to the error estimate for the 3.5σ trimming limit random error estimate. Using the 2.8σ trimming limits will typically reduce the error estimates relative to the 3.5σ values by 7 to 8%. Wind speed error estimates are, however, only reduced by 4%, as only wind speeds biased high are typically removed by quality assurance and the variability in wind speed is large relative to

Table I. The effect of different ICOADS trimming limits on the random error estimates for each basic variable

Variable	Ratio $2.8\sigma/3.5\sigma$	Ratio $4.5\sigma/3.5\sigma$
SST	0.92 ± 0.01	1.07 ± 0.01
Air temperature	0.92 ± 0.01	1.08 ± 0.01
Specific humidity	0.92 ± 0.01	1.06 ± 0.02
Surface pressure	0.93 ± 0.01	1.07 ± 0.01
Wind speed	0.96 ± 0.01	1.04 ± 0.004

its mean value. Using the 4.5σ trimming limits will typically increase the error estimates by 6 to 8%, but again the wind speed shows a smaller change of 4%.

5. DISCUSSION

5.1. Sources of random observational error

All observation types are subject to 'reporting preference'. This is the tendency for a human observer to round observations to a whole number (then, with decreasing preference: 0.5; 0.2 and 0.8; 0.6 and 0.4; 0.1, 0.3, 0.7 and 0.9) and is described in detail for SST by Kent and Taylor (in press). This reporting preference, whereby about half of the observations (and much more for some data sources) are reported in whole numbers, degrades the quality of the dataset but actually has little effect on the random error estimates.

Errors in VOS pressure reports derive from a number of sources. Mercury barometers need to be corrected for temperature, variations in gravity, 'index error' (essentially a calibration for the instrument), height and capillarity (Met Office, 1969). They can be subject to pumping errors due to wave motion. Aneroid barometers usually only require correction for height and index error, although some types do require temperature correction (Met Office, 1995). Knocking the aneroid barometer may alter its index correction, and aneroid barometers should not be exposed to rapid changes in temperature or direct sun. Dial aneroid barometers require tapping before the reading is taken, as the pointer is liable to stick. The index correction of dial aneroid barometers can be changed with an adjusting screw, and any changes need to be documented. Coding errors may occur, as the pressure is reported in millibars and tenths (equivalent to hectopascals and tenths) with the thousands figure omitted (for example, 1013.2 hPa is reported as 0132). On modern ships, the use of air conditioning and sealed bridges means that an external connection is required for the correct pressure to be recorded; it is not known how often this practice is followed. All barometers need regular calibration to maintain data quality. The wind speed may give rise to errors in pressure observations. The direct effect of the wind can be to increase the pressure, but effects are directional and hard to predict. The indirect effect of wind is a change of external pressure around the bridge, which should decrease the pressure with increasing wind speed. All of these factors may add to the variability of VOS pressure observations.

Coding errors can occur for wind speed, as there is an indicator flag giving whether the observation is in knots or metres per second and whether the method of observation is visual or using an anemometer. Unless the report is derived from a logbook then the reported wind speed will be in whole knots (or, less often, whole metres per second). Anemometer-derived winds can contain errors due to poor calibration and exposure, air flow distortion (Yelland *et al*., 2002; Moat *et al*., 2005), pumping due to wave action, and lack of sensitivity in low winds. The true wind speed needs to be derived from the relative wind speed and direction, the ship speed and direction, and any deviation of the ship's heading from the ship's direction of travel. This calculation has led to a significant number of incorrect wind speeds (Kent *et al*., 1993; Gulev, 1999; Smith *et al*., 1999). The occurrence of coding errors should have been reduced by the introduction of automatic coding software, such as TurboWin, SEAS and OBSJMA. Errors in direction due to any 'dead-band' in the direction can lead to incorrect true wind speeds being calculated. Visual wind speeds will contain random errors due to the imprecise method of relating wind speed to sea state. Visual wind observations are typically reported as the centre or the extremes of the Beaufort Scale interval, making the distribution of observations less uniform than those from anemometers. Atmospheric conditions and the presence of swell may also contribute to errors in visual observations, which are particularly difficult to make at night or in poor visibility. Visual winds suffer from reporting preference more than other observations, as the Beaufort Scale has relatively large graduations. Errors in wind speed are discussed in more detail by Thomas *et al.* (2005).

Although the use of automatic coding software has reduced coding errors, a recent source of both random error and bias into VOS wind reports was due to changes in the TurboWin logging software (details can be found on the KNMI website, http://www.knmi.nl/onderzk/applied/turbowin/turbowin.html, accessed 28 July 2004). This resulted in some wind speeds being corrected to 10 m height before transmission (as recommended by the WMO) and some were not (which had been common practice). There were no metadata to distinguish which reports had been corrected and which had not, which made the height correction of data using Publication 47 metadata problematic. This is of concern for estimates of wind speed trends, as the reported wind speed will be reduced.

Errors can be introduced into air temperature observations through poor exposure of the instruments, radiative heating of the sensor and ship (Berry *et al.*, 2004), proximity to heating sources (such as the engine room or galley), or heating by lights used to read the sensors. Errors will tend to be larger when relative wind speeds are light (Met Office, 1995) and ventilation rates, therefore, are low. Unshielded psychrometers used in the sun may contain errors, as they are not shielded from solar radiation (Met Office, 1995).

Humidity errors can derive from a number of sources (e.g. Met Office, 1995). Errors in the wet-bulb thermometer itself require index correction. If the wet bulb is covered with a layer of material that is too thick (or with ice), then the time constant of the thermometer is increased and this could lead to error. If the wick is contaminated with salt or other impurities then the wet-bulb temperature could be elevated (although this effect could not be detected in a sample of wicks collected from UK VOS, Peter Taylor, personal communication). At low temperatures there may be uncertainty as to whether the wet bulb is covered with ice, water, or a mixture. If the wick is badly designed, then stem heat conduction may occur. A poorly ventilated screen will also lead to increased random errors. If the wick is allowed to dry out, then the humidity will be overestimated. Errors in the dry-bulb temperature can also lead to humidity errors; these could be caused by inadequate ventilation or from moisture, such as spray, on the dry bulb, but not from the heating errors themselves (Kent and Taylor, 1996).

Kent and Taylor (in press) summarize the main causes of variability in SST observations, which include the effect of heat exchange with the atmosphere on bucket measurements, the use of different bucket designs and measurement instructions by observers on ships recruited by different countries, the effect of different measurement depths on engine intake and hull sensor measurements, and possible heating by the ship's engines on engine intake observations.

5.2. Summary comparison with error estimates of Kent et al. (1999)

Kent *et al.* (1999) made comparisons of their error estimates with those from the literature (Gilhousen, 1987; Weare, 1989; Wilkerson and Earle, 1990; Gleckler and Weare, 1987). Here, we compare our estimates of random observational error with those of Kent *et al.* (1999). Table II shows the mean estimates from the present study and those from the earlier Kent *et al.* (1999) study. Comparisons for the individual variables are described in the appropriate subsections of Section 4.

5.3. Comparison with estimates of total observational error

We have estimated the total observational error in the ICOADS 2° monthly summaries (Brohan *et al.*, 2003; Rayner *et al.*, 2005). It should be noted that for these calculations we have used a longer time period (1950 to 2002) than for the random measurement error estimates. Figure 2 shows that, over the period 1970 to 2002, the estimates of random measurement error have decreased. The contribution of measurement error to the total observational error is, therefore, likely to be greater for the period 1950 to 2002 than for 1970 to 2002. The contribution of measurement error to the total observational error may, therefore, be underestimated in the comparisons shown in this section. Figure 3 shows the variation with latitude of the estimated random

Table II. Global estimates of random observational errors from Kent *et al*. (1999) and from the present study (as in Figures 1 and 2). Uncertainties quoted are the standard deviation of the individual monthly 30° area estimates used to calculate the mean error estimate

Variable	Mean error		
	Kent *et al*. (1999)	Weighted by number of observations	Weighted by ocean area
Pressure (hPa)	2.3 ± 0.2	2.2 ± 0.4	2.1 ± 0.2
Wind speed (m s^{-1})	2.4 ± 0.2	2.3 ± 0.1	2.2 ± 0.1
10 m neutral wind speed (m s^{-1})	2.1 ± 0.2	2.0 ± 0.1	2.0 ± 0.1
Air temperature (°C)	–	1.4 ± 0.1	1.2 ± 0.1
10 m neutral air temperature (°C)	1.4 ± 0.1	1.1 ± 0.1	1.0 ± 0.1
Specific humidity (g kg^{-1})	–	1.1 ± 0.1	1.2 ± 0.1
10 m neutral specific humidity (g kg^{-1})	1.1 ± 0.2	1.0 ± 0.1	1.1 ± 0.1
SST (°C)	1.2 ± 0.1	1.3 ± 0.1	1.2 ± 0.1

measurement error and the total observational uncertainty. For pressure in the extratropics (Figure 3(a)) the measurement error is small compared with the total observational error. In the tropics, where pressure variability is small, the measurement error makes a large contribution to the total uncertainty. In high latitudes, where the pressure variability is large, the magnitude of the random observational error is one-fifth of the total observational error estimate. For the construction of monthly 2° area gridded pressure datasets, this shows that over most of the ocean the number of observations is much more important than their accuracy (although gross errors are always a problem). In the tropics, a smaller number of observations is required as long as those observations are of good quality. The requirements for other purposes, such as numerical weather prediction or the construction of datasets with different resolution, will be different to those presented here.

Wind speed error estimates show similar, but less extreme, variability with latitude, with the total observational uncertainty being about twice the magnitude of the estimated measurement error. In the highest latitudes the total uncertainty is just under three times the magnitude of the measurement uncertainty (Figure 3(b)). Similar variability is seen for air temperature (Figure 3(c)), although the total observational uncertainty estimates are much lower in southern high latitudes than in northern high latitudes. This may be a result of lower air temperature variability in this region, but confidence in both types of uncertainty estimate in the Southern Ocean is low.

Specific humidity total observational uncertainty is highest in mid latitudes, but the contribution of measurement error is greatest in the tropics (Figure 3(d)). It is for SST that the benefit of reducing measurement error is the most important (Figure 3(e)). In the tropics the measurement error estimates are slightly over half the magnitude of the total observational error estimates. In mid latitudes this typically reduces to about half the magnitude of the total observational error estimates. In high latitudes and in some limited regions, e.g. like the Gulf Stream area, the sampling uncertainty will be much higher.

6. CONCLUSIONS

Estimates of the random measurement error in VOS surface meteorological reports from ICOADS have been made for the period 1970 to 2002 for 30° areas. We have used the semivariogram method to estimate the random errors by attempting to separate the spatial and random components in variability (Lindau, 1995, 2003; Kent *et al*., 1999; Kent and Challenor, 2005).

Typically, the estimated measurement error decreases over the period 1970 to 2002, showing that VOS reports have improved in quality. Weighted by the number of VOS observations, the error estimates for VOS observations as reported are: pressure 2.2 ± 0.4 hPa; wind speed 2.3 ± 0.1 m s^{-1}; air temperature 1.4 ± 0.1 °C; specific humidity 1.1 ± 0.1 g kg^{-1}; SST 1.3 ± 0.1 °C. VOS observations of wind speed, air

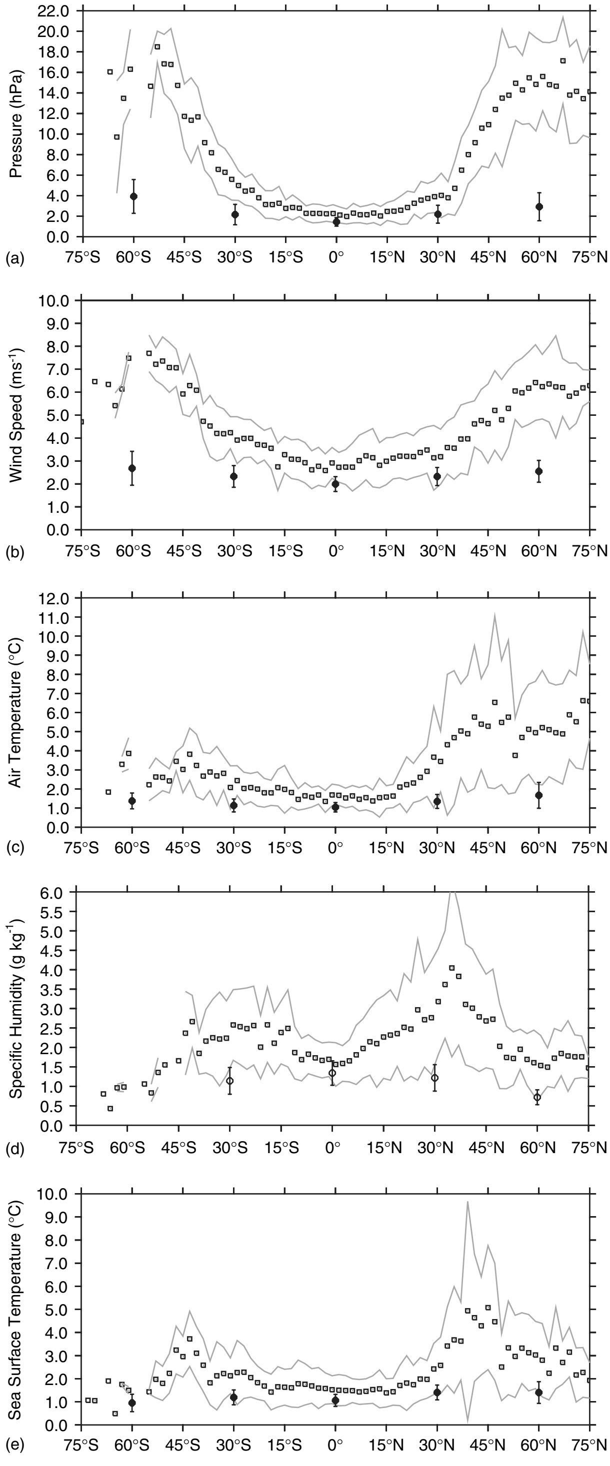

Figure 3. Comparison of random measurement error estimates (as in Figures 1 and 2, circles) with estimates of total observational (measurement plus sampling, squares) error estimates. Error bars and ranges shown are one standard deviation of the mean: (a) pressure (hPa); (b) wind speed (m s^{-1}); (c) air temperature (°C); (d) specific humidity (g kg^{-1}); (e) SST (°C)

temperature and humidity can be adjusted to a common reference measurement height of 10 m using metadata on measurement heights from WMO Publication No. 47 and to neutral atmospheric stability. If this is done, then the random measurement errors are reduced for these variables: 10 m neutral wind speed 2.0 ± 0.1 m s^{-1}; 10 m neutral air temperature $1.1 \pm 0.1\,^{\circ}$C; 10 m neutral specific humidity 1.0 ± 0.1 g kg^{-1}. The reduction in random error illustrates the worth of making these corrections for height and stability, but their main purpose is to reduce bias in the observations, which is not detectable by the method used in this study.

Estimates of measurement error have been compared with estimates of the total measurement plus sampling error calculated for ICOADS 2° monthly summaries (Brohan *et al*., 2003; Rayner *et al*., 2005). In regions of high variability the sampling error dominates the measurement error, especially for pressure and wind speed in high latitudes. For all variables in the tropics the measurement error contributes half or more of the total error, and for air temperature, specific humidity and SST it makes a substantial contribution to the total error at all latitudes.

ACKNOWLEDGEMENTS

This work was supported by funding from the UK Government Meteorological Research Programme. We thank Philip Brohan of the Met Office and Peter Challenor of the Southampton Oceanography Centre for advice on the calculation of the estimates of total observational error. ICOADS was supplied by Steven Worley of the National Center for Atmospheric Research Data Support Section. WMO Report No. 47 ASCII files were provided by Joe Elms at the National Climatic Data Center, Asheville, NC, and the WMO (ftp://www.wmo.ch/wmo-ddbs/, accessed 28 July 2004). Scott Woodruff of the NOAA Climate Diagnostics Center provided help and advice. The Ferret program of NOAA's Pacific Marine Environmental Laboratory (http://www.ferret.noaa.gov/, accessed 28 July 2004) was used for analysis and graphics in this paper. We also thank the reviewers for their comments, which have helped us to improve the paper.

REFERENCES

Berry DI, Kent EC, Taylor PK. 2004. An analytical model of heating errors in marine air temperatures from ships. *Journal of Atmospheric and Oceanic Technology* **21**(8): 1198–1215.

Brohan P, Rayner NA, Vanicek M, Tett SFB, Kent EC. 2003. Combined estimates of uncertainties in gridded marine temperature fields due to measurement errors and under-sampling of variability. In *Proceedings of Second JCOMM Workshop on Advances in Marine Climatology (CLIMAR-II)*, Brussels, 17–22 November. WMO/TD-No. 1199 (available on CD-ROM from World Meteorological Organization, Geneva, Switzerland).

Diaz H, Folland C, Manabe T, Parker D, Reynolds R, Woodruff S. 2002. Workshop on advances in the use of historical marine climate data. *WMO Bulletin* **51**: 377–380.

Gilhousen DB. 1987. A field evaluation of NDBC moored buoy winds. *Journal of Atmospheric and Oceanic Technology* **4**(1): 94–104.

Gleckler PJ, Weare BC. 1997. Uncertainties in global ocean surface heat flux climatologies derived from ship observations. *Journal of Climate* **10**: 2764–2781.

Gulev SK. 1999. Comparison of COADS release la winds with instrumental measurements in the northwest Atlantic. *Journal of Atmospheric and Oceanic Technology* **16**(1): 133–145.

Gulev SK, Grigorieva V, Sterl A, Woolf DK. 2003. Assessment of the reliability of wave observations from voluntary observing ships: insights from the validation of a global wind wave climatology based on voluntary observing ship data. *Journal of Geophysical Research* **108**(C7): 3236. DOI: 10.1029/2002JC001437.

Hsiung J. 1986. Mean surface energy fluxes over the global ocean. *Journal of Geophysical Research* **91**(C9): 10 585–10 606.

Isemer H-J, Hasse L. 1987. *The Bunker Climate Atlas of the North Atlantic Ocean. Vol. 2: Air–Sea Interactions*. Springer-Verlag.

Jones PD, Osborn TJ, Briffa KR. 1997. Estimating sampling errors in large-scale temperature averages. *Journal of Climate* **10**: 2548–2568.

Josey SA, Kent EC, Taylor PK. 1999. New insights into the ocean heat budget closure problem from analysis of the SOC air–sea flux climatology. *Journal of Climate* **12**(9): 2856–2880.

Kent EC, Challenor PG. In press. Towards estimating climatic trends in SST data, part 2: random errors. *Journal of Atmospheric and Oceanic Technology*.

Kent EC, Kaplan A. Accepted for publication. Towards estimating climatic trends in SST data, part 3: systematic errors. *Journal of Atmospheric and Oceanic Technology*.

Kent EC, Taylor PK. 1996. Accuracy of humidity measurements on ships: consideration of solar radiation effects. *Journal of Atmospheric and Oceanic Technology* **13**(6): 1317–1321.

Kent EC, Taylor PK. 1997. Choice of a Beaufort Equivalent Scale. *Journal of Atmospheric and Oceanic Technology* **14**(2): 228–242.

Kent EC, Taylor PK. In press. Towards estimating climatic trends in SST data, part 1: methods of measurement. *Journal of Atmospheric and Oceanic Technology*.

Kent EC, Taylor PK, Truscott BS, Hopkins JS. 1993. The accuracy of voluntary observing ship's meteorological observations — results of the VSOP-NA. *Journal of Atmospheric and Oceanic Technology* **10**(4): 591–608.

Kent EC, Taylor PK, Challenor PG. 1998. A comparison of ship and scatterometer-derived wind speed data in open ocean and coastal areas. *International Journal of Remote Sensing* **19**(17): 3361–3381.

Kent EC, Challenor PG, Taylor PK. 1999. A statistical determination of the random observational errors present in voluntary observing ships meteorological reports. *Journal of Atmospheric and Oceanic Technology* **16**: 905–914.

Lindau R. 1995. A New Beaufort Equivalent Scale. In *Proceedings of the International COADS Winds Workshop*, Kiel, Germany, Institut für Meereskunde and NOAA Environmental Research Labs; 232–252.

Lindau R. 2003. Errors of Atlantic air–sea fluxes derived from ship observations. *Journal of Climate* **16**: 783–788.

Met Office. 1969. *Marine Observer's Guide*, ninth edition. HMSO: London.

Met Office. 1995. *Marine Observer's Guide*, eleventh edition. HMSO: London.

Moat BI, Yelland MJ, Pascal RW, Molland AF. 2005. An overview of the airflow distortion at anemometer sites on ships. *International Journal of Climatology* **25**: 997–1006; this issue.

Morone LL. 1986. The observational error of automated wind reports from aircraft. *Bulletin of the American Meteorological Society* **67**: 177–185.

National Geophysical Data Center. 1998. Digital relief of the surface of the earth, data announcement 88-MGG-02, NOAA, Boulder, Colorado, USA. http://www.ngdc.noaa.gov/mgg/ [accessed 21 March 2005].

Oberhuber JM. 1988. An atlas based on the 'COADS' data set: the budgets of heat, buoyancy and turbulent kinetic energy at the surface of the global ocean. Max-Planck-Institut für Meteorologie, Report No 15, Hamburg, 20pp plus figures.

Parker DE, Kent EC, Woodruff SD, Dehenauw D, Harrison DE, Manabe T, Mietus M, Swail VR, Worley SJ. 2004. The Second JCOMM Workshop on Advances in Marine Climatology (CLIMAR-II). *WMO Bulletin* **53**(2): 157–159.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global analyses of SST, sea ice and night marine air temperature since the late 19th century. *Journal of Geophysical Research* **108**(D14): 4407. DOI: 10.1029/2002JD002670.

Rayner NA, Brohan P, Parker DE, Folland CK, Kennedy J, Vanicek M, Ansell T, Tett SFB. 2005. Improved analyses of change and uncertainties in sea surface temperature measured *in situ* since the mid-nineteenth century. *Journal of Climate*, submitted for publication.

Smith SR, Bourassa MA, Sharp RJ. 1999. Establishing more truth in true winds. *Journal of Atmospheric and Oceanic Technology* **16**(7): 939–952.

Smith TM, Reynolds RW. 2004. Improved extended reconstruction of SST (1854–1997). *Journal of Climate* **17**(12): 2466–2477.

Thomas BR, Kent EC, Swail VR. 2005. Methods to homogenize wind speeds from ships and buoys. *International Journal of Climatology* **25**: 979–995; this issue.

Weare BC. 1989. Uncertainties in estimates of surface heat fluxes derived from marine reports over the tropical and subtropical oceans. Tellus, *Series A: Dynamic Meteorology and Oceanography* **41**: 357–370.

Wilkerson JC, Earle MD. 1990. A study of differences between environmental reports by ships in the voluntary observing program and measurements from NOAA buoys. *Journal of Geophysical Research* **95**(C3): 3373–3385.

WMO. 1994. International list of selected, supplementary and auxiliary ships. WMO Report No. 47. WMO, Geneva.

Wolter K. 1997. Trimming problems and remedies in COADS. *Journal of Climate* **10**: 1980–1997.

Woodruff SD, Slutz RJ, Jenne RL, Steurer PM. 1987. A comprehensive ocean–atmosphere data set. *Bulletin of the American Meteorological Society* **68**: 1239–1250.

Woodruff SD, Lubker SJ, Wolter K, Worley SJ, Elms JD. 1993. Comprehensive Ocean–Atmosphere Data Set (COADS) release 1a: 1980–92. *Earth System Monitor* **4**(1): 4–8.

Woodruff SD, Diaz HF, Elms JD, Worley SJ. 1998. COADS release 2 data and metadata enhancements for improvements of marine surface flux fields. *Physics and Chemistry of the Earth* **23**: 517–527.

Yelland MJ, Moat BI, Pascal RW, Berry DI. 2002. CFD model estimates of the airflow over research ships and the impact on momentum flux measurements. *Journal of Atmospheric and Oceanic Technology* **19**(10): 1477–1499.

